

УДК 004.738.5:004.451.53:004.22

## СЕМАНТИЧКИ ВЕБ И ЕЛЕКТРОНСКИ ИЗВОРИ ИНФОРМАЦИЈА\*

Даница Радовановић  
Факултет за менаџмент малих предузећа

### Сажетак

Коришћење електронских извора информација зависи од добрих могућности претраживања, а појам Семантичког веба (Semantic Web) је настао у трагању за ефикаснијим решењима за проналажење информација. World Wide Web (WWW), уз помоћ машина за претраживање и огромног броја расположивих метаподатака омогућава добијање информација али које тек у неком степену могу задовољити информациону потребу корисника. Истовремено је све више истраживачких напора да се на самом Вебу повећа ефикасност претраживања до добијања што релевантнијих информација. Као један од најновијих резултата таквих напора W3C конзорцијума, Семантички веб представља скуп организованих технолошких стандарда, производа и информација повезаних на такав начин да могу бити индексиране и семантички филтриране кроз процес обраде на глобалном нивоу. Семантички веб доприноси ефикасном претраживању и тиме што омогућава посебан начин представљања информација на Вебу које се могу посматрати и као глобално повезане онлајн базе података. Архитектуру Семантичког веба чине три важна нова стандарда - XML (eXtensible Markup Language), RDF (Resource Description Framework) и онтологије. Семантички веб је у развоју и још увек није у општој употреби, али обећава радикално побољшање могућности проналажења, сортирања и класификовања информација.

**Кључне речи:** Семантички веб, електронски извори информација, проналажење информација, представљање информација, Интернет, стандарди

## SEMANTIC WEB AND ELECTRONIC INFORMATION RESOURCES

Danica Radovanović  
Faculty for Management of Small Enterprises

### Abstract

The usage of electronic resources depends on good possibilities of searching and concept of the Semantic Web can be convenient solution for information retrieval (IR). WWW (World Wide Web) enables, with help of the search engines and huge number of available (meta)information, data that can satisfy user information need, but only at some extent. At the same time, there are more and more research efforts to increase the efficiency for IR until one gets as much as relevant information on the Web. As one of the latest results of this W3C efforts, Semantic Web presents a group of organized technological standards, IT products, and information linked in such a way that can be easily indexed and semantically filtrated through process of classification on global level. Semantic Web and its principles make IR easier because it can be also observed as very useful and successful way of representing data on WWW or as a group of globally linked databases. The architecture of Semantic Web consists of three important IT standards: XML (eXtensible MarkUp Language), RDF (Resource Description Framework) and the ontologies. Semantic web is still under development and is not in common usage but it promises that it will radically improve the possibility of searching, sorting and classification of information.

**Key words:** Semantic Web, electronic information resources, information retrieval, information representation, Internet, standards

## 1. ИНФОРМАЦИОНЕ ПОТРЕБЕ, WORLD WIDE WEB И СЕМАНТИЧКИ ВЕБ

Као у многим областима глобалних активности, информационо - комуникационе технологије су одавно нашле примену у библиотекаству, пре свега у претраживању информација и коришћењу база података на Интернету. Коришћење електронских извора информација зависи од добрих могућности

---

\* Рад је изложен на Стручном скупу *Коришћење електронских извора информација*, одржаном у оквиру Осме скупштине Заједнице библиотека универзитета у Србији, 26. септембра 2002. године у Народној библиотеци Србије у Београду.

претраживања а концепт Семантичког веба (Semantic Web) је настао у трагању за ефикаснијим решењима за проналажење информација.

С обзиром да већ више од десет година постоји World Wide Web (WWW), уз помоћ машина за претраживање (search engines) огроман број расположивих метаподатака омогућава добијање информација који тек у неком степену могу задовољити информациону потребу. Истовремено је све више истраживачких напора да се на самом Вебу повећа ефикасност претраживања до добијања што релевантнијих информација. Потреба за таквим напорима опстаје јер се дешава да су резултати претраживања који се добијају, често потпуно ирелевантни, те је корисник неког информационог система или базе принуђен да користи побољшавања упита и, често, накнадне мануелне операције филтрирања да би дошао до прецизнијих резултата.

Tim Berners-Lee, професор на Massachusetts Institute of Technology (MIT), који је 1989. године развио World Wide Web креирајући језик за представљање и повезивање садржаја и узајамну размену информација, 1994. године основао је и World Wide Web Consortium (W3C) на поменутом Институту (<http://www.w3.org>). У даљем развоју овог сервиса на Интернету, заједно са својим колегама, Berners-Lee ради на његовој семантици тако да извори информација на њему неће бити само лоцирани већ ће, уз помоћ рачунара, бити и семантички обрађени. Као један од најновијих резултата таквих напора W3C конзорцијума, Семантички веб представља скуп организованих технолошких стандарда, производа, и информација повезаних на такав начин да могу бити лако доступне и семантички филтриране кроз процес обраде на глобалном нивоу.

## 2. ДЕФИНИЦИЈЕ

Семантички веб је апстрактно представљање података на World Wide Web-у, а заснован је на окружењу **RDF** (Resource Description Framework) и другим стандардима. У литератури се за овај концепт среће више дефиниција.

'Семантички веб је наставак, проширење постојећег Веба на којем је информацији дато прецизно дефинисано значење и који боље омогућава сарадњу између рачунара и корисника.' - *Tim Berners-Lee, James Hendler, Ora Lassila, [The Semantic Web](#), Scientific American, May 2001.*

'Семантички веб је Веб сачињен од података који могу бити процесирани директно или индиректно уз помоћ машина' - *Tim Berners-Lee у [Weaving the Web](#).*

Семантички веб као концепт омогућава да доступни извори информација на Вебу могу бити организовани и коришћени не само синтаксним или структурним методама, већ и семантичким. Он представља синергију програмских апликација који прикупљају садржај са Веба из различитих извора, затим процесирају информацију и размењује резултате са другим програмима.

Семантички веб доприноси ефикасном претраживању и тиме што омогућава посебан начин представљања информација које се на Вебу могу посматрати и као скупови глобално повезаних онлајн база података. Да би Семантички веб функционисао рачунари треба да имају приступ структурисаним колекцијама информација и да утврде дефинисана правила аутоматизованог управљања. Машине за претраживање на Интернету за решавање проблема недостатка значења на Вебу и анализу презентација, користе кључне речи и појмове који се подударaju са терминима у упитима које поставља корисник. Док су на почетку радиле са кључним речима, следећа генерација машина за претраживање је испитивала контекст аутоматизованом или мануелном провером веза (линкова) између страница и садржаја на Вебу, превodeћи кључне речи и сажетке у појмове. Аутоматизована класификација и категоризација резултата је сада уобичајена; савремена технологија претраживања обухвата технологију која проналази тражене атрибуте у Веб страницама и повезује их са профилима који описују карактеристике резултата за корисника.

## 3. АРХИТЕКТУРА СЕМАНТИЧКОГ ВЕБА

Док **HTML** (Hypertext Markup Language) служи за представљање података и њихов изглед на Вебу и форматирањем тагова описује како изгледа информација на Веб страници, архитектуру семантичког Веба чине два важна информационо технолошка стандарда, као и трећи који носи кључну улогу:

- **XML** (eXtensible Markup Language) - прошириви језик за означавање који одређује структуру података,
- **RDF** (Resource Description Framework) - централни протокол на Вебу заснован је као W3C стандард када су у питању семантичке везе. RDF, заправо, описује семантичке везе између електронских извора.
- и – **онтологије**, најбитнији и најтеже остварив чинилац Семантичког Веба.

### 3.1 XML (eXtensible Markup Language)

XML је добро познат W3C стандард из групе језика [SGML](#) ("[Standard Generalized Markup Language](#)"), настао 1996. године, који је служи за опис и размену података на Вебу. XML омогућава ауторима да креирају сопствено означавање (markup), синтаксу (нпр. <Autor>) који носи у себи део семантике. Таг, као што је <Autor>, носи више семантике него таг <H1>. Рачунар једноставно не зна шта је "аутор" и у каквом је односу појам "аутор", на пример, са појмом "особа". XML може да помогне креаторима, кориснику да предвиди која информација може стајати "између тагова". XML омогућава дизајнерима презентација на Вебу да спецификују своје сопствене тагове у оквиру докумената за Веб. Тагови постављају информацију у структуралну хијерархију тако да их корисник може разумети. Постојећи тагови у HTML страницама кодирају углавном једноставне ствари као што је изглед – layout документа (подвучено, *италик*, **болд** итд.) или хипертекстуалне везе (линкове), али је за креирање семантичког Веба потребан велики број тагова. Са њима треба да се опише предмет из текста и ко је аутор, тако да би машина за претраживање могла да разликује, на пример, дела која је написао Кафка, или есеј, чланак о аутору Кафки. Да би (међу)везе биле потпуне, сви елементи морају до детаља бити спецификовани тако да сам XML није довољан за реализацију семантике.

### 3.2 RDF (Resource Description Framework)

**RDF** (Resource Description Framework), као централни протокол на Вебу и заснован је као W3C стандард када су у питању семантичке везе - [www.w3.org/RDF](http://www.w3.org/RDF). Он, заправо, описује **семантичке везе** између електронских извора. RDF интегрише различите апликације - од библиотечких каталога и светских водича, портала, преко обједињених и скупљених података, софтвера и садржаја, до мултимедијалних колекција (музика, фотографија, записа догађаја), каталога - користећи XML као синтаксу за међусобну, узајамну размену. RDF спецификације омогућавају онтолошки систем који подржава информације на Вебу описујући семантику структуре у триплету: субјекат, предикат, објекат. Ови триплети се могу написати уз помоћ XML тагова (видети шему RDF примера у XML-у). Уз помоћ RDF стандарда, документ је класификован тако да одређени објекти (људи, садржаји на Вебу и слично) имају својства којима се могу додељивати одређене вредности. RDF триплети стоје између информација на Семантичком вебу, омогућавајући рачунарима да их лоцирају, прикупе тражене информације и имају увид у њихово кретање.

На тај начин представљени модел садржи податке о ресурсима, изворима. Извор подразумева све што је повезано са **URI** (Universal Resource Identifier). Субјект и објект су сваки понаособ идентификовани Универзалним идентификатором извора - (URI), који се користи у повезивању Веб страница док су URL-ови - Uniform Resource Locators, најпознатија врста URI-ија). Глаголи су такође спецификовани универзалним идентификатором извора, што омогућава да се утврди нови концепт, нови глагол, тако што се дефинише URI за сваки од њих негде на Вебу. Поред универзалних идентификатора извора, који су препознатљиви глобално, важни су и концепти: јединствене дефиниције елемената, системи са одређеним правилима процесирања, и уобичајени протоколи инфраструктуре Интернета. Све врсте метаподатака (metadata) морају да користе утврђене кључне речи (key words), у супротном претраживач не може да функционише када корисник постави упит. Следи конкретан пример употребе елемената стандарда Dublin Core за метаподатке уз примену RDF-а (многе базе података на Интернету и библиотечки узајамни каталози у свету користе овај концепт).

Такође, важни су и односи између чињеница, података, везе које постоје међу њима, као и подаци који су релевантни за одређени упит или захтев. Следи пример веза које упућују са једног URL - једне локације на другу (Веб страницу). ID-идентитет RDF извора је ID једне од више предметних тачака на чворовима мапе.

### <RDF primer u XML-u>

```

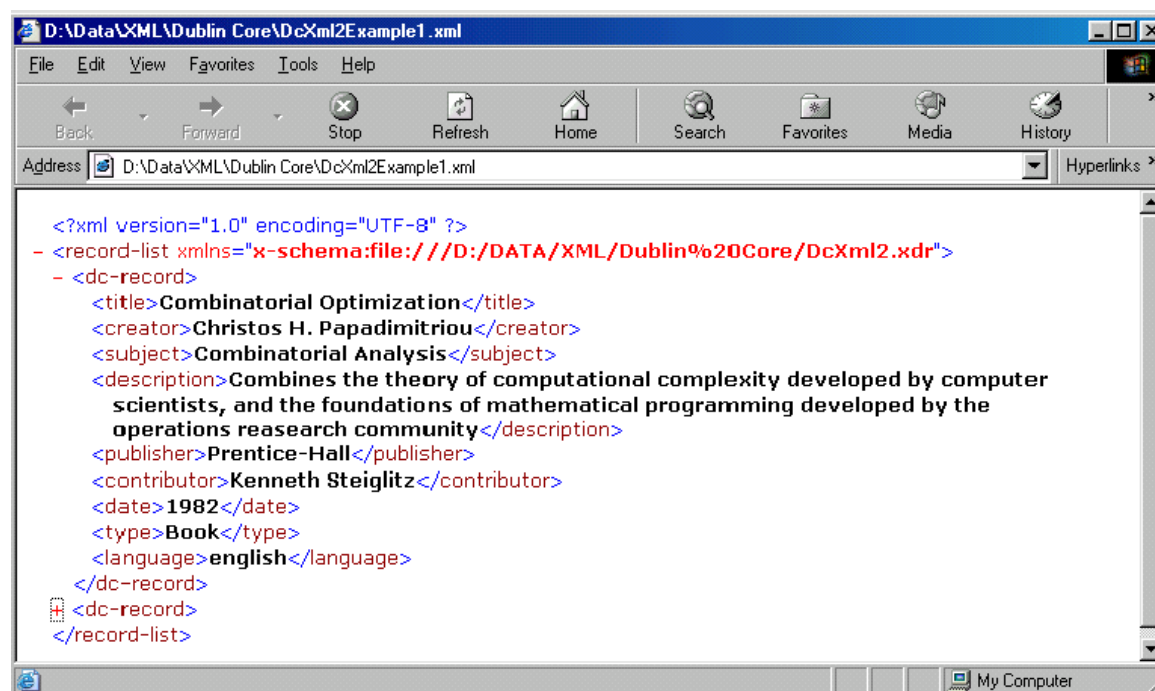
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.ilrt.org/people/cmdjb/">
    <dc:title>Dave Beckett's Home Page</dc:title>
    <dc:creator>Dave Beckett</dc:creator>
    <dc:publisher>ILRT, University of Bristol</dc:publisher>
  </rdf:Description>
</rdf:RDF>

```

Циљ Семантичког веба је да омогући софтверу да пронађе тражени податак на Вебу, да га семантички разуме, направи везу-упутницу, тј. трансфер и све примени на одговарајући захтев, упит. То подразумева комбинацију XML-а, RDF-а и концепта који се зове онтологија.

### 3.3 Онтологије као механизам спецификација

Семантички веб не може да се реализује без метаподатака (метадата). Метаподатак омогућава посреднички, међуниво у претраживању тако да корисник не претражује масивне количине ирелевантног пуног HTML текста који се налази на Вебу. Метаподаци у Веб страницама или каталозима морају бити повезани са специјалним документима који дефинишу термине метаподатака и односе између термина. Ови скупови појмова и њихових међувеза се називају "онтологијама".



```

<rdf:RDF xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdf="http://www.w3c.org/2000/01/rdf-schema#"
  xmlns:tms="http://www-db.stanford.edu/rdf-tmmapping/tm-schema#"
  xmlns="http://www-db.stanford.edu/rdf-tmmapping/tm-schema#">
  <rdfs:Class ID="t"/>
  <rdfs:Class ID="a"/>
  <rdfs:Class ID="s"/>
  <rdf:Property ID="associationMember"/>
  <rdf:Property ID="associationScope"/>
  <rdf:Property ID="associationTemplate"/>
  <rdf:Property ID="scopeComponent"/>
  <rdf:Property ID="roleLabel"/>
  <rdf:Property ID="scr"/>
  <rdf:Property ID="sir"/>
</rdf:RDF>

```

За регистровање односа међу објектима и њихових карактеристика, конструктори Семантичког веба користе **онтологије** – које су трећа основна компонента Семантичког веба. Сам термин онтологија, преузет из филозофије, а део је метафизике о природи постојања - *per se*. У информационом контексту коришћења електронских извора и знања, онтологија се дефинише као спецификација идеја, односно, концептуализације. Концептуализација је апстраховано представљање света или неког његовог дела, са неког становишта и из неког одређеног разлога. Свака информациона база или база заснована на знању је повезана са неком концептуализацијом, експлицитно или имплицитно.

Онтологија је експлицитна спецификација концептуализације којом се дефинишу термини, појмови и њихове међусобне везе. Онтологија је опис (као формална спецификација програма) појмова и веза које могу постојати унутар хијерархије у систему појмова. Очигледно је да је велики задатак да се изграде такве онтологије које ће описати системе појмова о свему што је представљено на Вебу. Онтологије могу наћи свој зачетак у таксономијама - деловима онтологија који представљају класификацију, структурни распоред података у оквиру класа и категоризују одређено подручје и везе између елемената. Резултат су студија именовања података, записа и предмета у генеричким схемама, као и правила који успостављају односе између података. Пример дигиталне таксономије: <http://www.geocities.com/RainForest/Vines/8695>.

Семантички веб описује онтологије у RDF стандарду користећи XML као интрумент. Онтологије омогућавају рачунару да се прикупе информације са различитих страна на Интернету и различитих Веб презентација, и даје одговоре на софистициране упите корисника. Укратко, семантички Веб претвара рачунар у личног асистента за претраживање који је способан да урони у читаву структуру информација и пронађе тражени податак брзином светлости. Библиотеке и информациони центри би могли да подрже овај пројекат зато што располажу знањем у грађењу оваквих врста механизма класификације.

#### 4. БУДУЋНОСТ СЕМАНТИЧКОГ ВЕБА И БИБЛИОТЕКЕ

Семантички веб је визија која је почела да се остварује: имати тражени податак на Вебу, дефинисан и повезан, линкован на начин, који може бити коришћен машинским путем - не само за сврху приказивања већ и за употребу у различитим апликацијама претраживања и доступности података. Технологија за реализацију семантичког Веба постоји и познат је начин како изградити терминологију и како користити метаподатке. Цела идеја зависи од коришћења заједничких стандарда.

Tim Berners-Lee је убеђен да ће се у блиској будућности Семантички веб користити тиме што се повећава примена RDF-у компатибилних докумената. Он семантички Веб види као могућност којом ће се разбити баријере (интелектуалне и културне) које данас постоје на Вебу. Претпоставља се да ће пуну снагу и примену семантички Веб достићи до 2010. године.

Библиотекари и информациони стручњаци могу да подрже овај пројекат с обзиром да поседују знања у грађењу механизма – онтологија. Семантички веб треба да одигра важну улогу за библиотеке и њихове кориснике, тим пре што ће претраживања уз помоћ Семантичког веба давати тачније одговоре чиме ће се бољи задовољити информационе потребе. Улога библиотекара би се огледала у стварању онтологија уз стварање сопствених база података и онлајн доступних каталога, нпр., у чувању записа у XML-у уз коришћење RDF стандарда, као и перманентном грађењу онтологија.

#### 4.1 Семантички Веб и пракса

Семантички веб би требало да дефинише структуру садржаја страница на Вебу и да креира услове под којима би програми - агенти "ишли" од странице до странице ради добијања одговора на упите корисника.

Постоје институције које се међу првима почеле да користе принципе Семантичког веба. DARPA (U.S. Defense Advanced Research Projects Agency) и комерцијалне установе као Network Inference у Manchester-у, Енглеска, одавно већ користе инструменте за развој инфраструктуре семантичког Веба као и апликације за његово коришћење. У овом тренутку постоје компаније и институти у којима се ради на њима, нпр. Nokia (Espoo, Финска) - Markup, Hewlett-Packard - Java алатке за креирање и одржавање метаподатака, MIT или Free University of Amsterdam, а има и више језика за израду онтологија и др.

#### 4.2 Примери примене Семантичког веба

Поред Bot – софтвера на Интернету који поседује карактеристике интелигентног софтвера (intelligent agents) да огромном брзином на нивоу дневних ажурирања посећује Веб презентације и индексира их и најуспешнијег, за сада, примера робота за индексирање (друга генерација) - [www.google.com](http://www.google.com), следе неки од примера који користе структуру Семантичког веба:

- Mat-net <http://www.math-net.de/> користи структуру Семантичког веба за свој портал (област математичких наука)
- Пример RiboWeb базе која функционише са RDF схемом <http://www.medwebplus.com/obj/25678>

### 5. ЗАКЉУЧАК

Семантички веб је још визија која подразумева да на Вебу постоји тражени податак, дефинисан и повезан (линкован) на начин да се може користити машинским путем – не само за сврху приказивања, већ и за употребу у различитим апликацијама.

Семантички веб ће се реализовати заједно са даљим развојем стандарда и информационих технологија (RDF, Z39.50, DC, FRBR, XML). Z39.50, на пример, је скоро интернационални стандард који омогућава размену информација, метаподатака чак између разнородних рачунарских система. У овом тренутку, међутим, још је под знаком питања да ли ће концепција Семантичког веба бити успешна. Информатичари заједно са библиотекарима раде на решавању проблема претраживања информација и размене знања и информација. Трошкови око креирања онтологија су велики: оне одузимају доста времена и изискују перманентно развијање. Према неким мишљењима цела замисао око Семантичког веба је контрадикторна.

Технологија за реализацију Семантичког веба постоји и познат је начин како изградити терминологију и како користити метаподатке. Идеја зависи од коришћења заједничких стандарда. Семантички веб је у развоју и још увек није у општој употреби, али има изгледа да ће радикално побољшати могућност проналажења, сортирања и класификовања информација.



## 6. ЛИТЕРАТУРА

### УРЛ референце

- The World Wide Web Consortium Semantic Web: <http://w3.org/2001/sw>
- <http://www.w3.org/2000/10/swap/Primer> (Getting Into RDF & Semantic Web Using N3)
- <http://www.w3.org/DesignIssues/Semantic> (Semantic Web Roadmap)
- <http://purl.org/swag/whatIsSW> (What Is The Semantic Web?)
- <http://uwimp.com/eo.htm> (Semantic Web Primer)
- <http://logicerror.com/semanticWeb-long> (Semantic Web Introduction - Long)
- <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html> (SciAm: The Semantic Web)
- <http://www.xml.com/pub/a/2001/03/07/buildingsw.html> (Building The Semantic Web)
- <http://infomesh.net/2001/06/swform/> (The Semantic Web, Taking Form)
- <http://www.w3.org/2001/sw/Activity> (SW Activity Statement)
- <http://www.w3.org/2000/01/sw/> (SWAD) Semantic web road map / Tim Berners-Lee - <http://www.w3.org/DesignIssues/Semantic.html>
- Web design issues; what a semantic can represent / Tim Berners-Lee [www.w3.org/DesignIssues/RDFnot.html](http://www.w3.org/DesignIssues/RDFnot.html)
- Realising the full potential of the web / Tim Berners-Lee <http://www.w3.org/1998/02/Potential.html>
- [http://h10014.www1.hp.com/buildguide/meta\\_data\\_standards.htm](http://h10014.www1.hp.com/buildguide/meta_data_standards.htm) HP презентација -metadata standards
- Берклијево електронско издање - Finding Information on Internet [www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html](http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html)

### Референце о онтологијама

- Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.), Formal Ontology in Information Systems. Proc. of the 1st International Conference, Trento, Italy, 6-8 June 1998. IOS Press  
<http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf>
- Ontology, Conceptual Modeling, and Knowledge Engineering Group at: Institute for Systems Science and Biomedical Engineering of the Italian National Research Council (CNR), Padova, Italy  
<http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html>
- **SENSUS** - 70,000-node terminology taxonomy. SENSUS је наставак и реорганизација WordNet –а (укључујући остале онтологије повезане пројектом ISI) <http://mozart.isi.edu:8003/sensus2/>
- Остали пројекти онтологија доступни су на:  
<http://www.isi.edu/natural-language/projects/ONTOLOGIES.html>